Deep Full-Body Motion Network for a Soft Wearable Motion Sensing Suit

Dooyoung Kim, Junghan Kwon, Seunghyun Han, Yong-Lae Park, *Member, IEEE*, and Sungho Jo, *Member, IEEE*

***Abstract*—Soft sensors are becoming more popular in wearables as a means of tracking human body motions due to their high stretchability and easy wearability. However, previous research not only was limited to only certain body parts, but also showed problems in both calibration and processing of the sensor signals, which are caused by the high nonlinearity and hysteresis of the soft materials and also by the misplacement and displacement of the sensors during motion. Although this problem can be alleviated through redundancy by employing an increased number of sensors, it will lay another burden of heavy processing and power consumption. Moreover, complete full-body motion tracking has not been achieved yet. Therefore, we propose use of deep learning for full-body motion sensing, which significantly increases efficiency in calibration of the soft sensor and estimation of the body motions. The sensing suit is made of stretchable fabric and contains 20 soft strain sensors distributed on both the upper and the lower extremities. Three athletic motions were tested with a human subject, and the proposed learning-based calibration and mapping method showed a higher accuracy than traditional methods that are mainly based on mathematical estimation, such as linear regression.**

***Index Terms*—Body motion tracking, deep learning, soft sensors, soft wearables.**

I. Introduction

HUMAN motion tracking has been widely used in various research and commercial applications, such as biomechanics study, rehabilitation, three-dimensional (3-D) animation for video games and movies, and virtual/augmented reality (VR/AR) [1]–[4].

One of the main requirements for motion tracking is to accurately estimate the position and orientation of the joints or the segments of the human body based on the measurements from the sensors. To meet this requirement, soft wearable sensors are becoming popular due to their high stretchability and easy wearability. Soft sensors can be either directly attached to the skin of the wearer or sewn on a garment to be worn, and track the motions of the specific body parts, such as ankles, knees, hips, shoulders, and hands, without restricting the natural degrees of freedom (DOF) of the wearer [5]–[9].

Although previous studies have shown feasibility of motion sensing of the human body using soft sensors, they have been mostly limited to certain areas, and full-body motion tracking using soft sensors has not been reported yet, to the best of our knowledge [5]–[7], [10]. In order to implement this, the following challenges should be addressed.

First, the soft sensors typically show high nonlinearity and hysteresis in response [11], [12]. Moreover, it is more difficult to estimate body motions from the sensor signals if multiple mechanical stimuli, such as strain and pressure are applied. Although the pressure effect can be minimized by carefully selecting the locations of the sensors [5], it makes the design process complicated. It has recently been proposed to solve the problem of nonlinearity and hysteresis of soft sensors using deep learning [13], but it was limited only to a single sensor.

Another challenge is calibration of a large number soft sensors integrated to the suit, which makes the calibration process time-consuming and complex. A human body has many joints with a single DOF or multiple DOFs and, the same as or more number of sensors than the number of DOFs in each joint are required. Although it has been proposed to calibrate the multiple sensors using linear regression (LR) [6], it was limited to a single joint.

Therefore, we propose a deep full-body motion network (DFM-Net) for tracking full-body motions using a soft wearable sensing suit (see Fig. 1) in this paper. The suit contains 20 soft microfluidic strain sensors distributed both on the upper and the lower bodies, and they are simultaneously calibrated and processed to estimate the 3-D body motions. A deep neural network for temporal sequence modeling was implemented to take care of nonlinearity and hysteretic responses of the soft sensors. Moreover, it was constructed to adequately represent the relationship between the motions from multi-DOF body joints and the sensing data. In order to verify our learning-based cali-

Manuscript received January 22, 2018; revised June 15, 2018; accepted September 30, 2018. Date of publication October 8, 2018; date of current version February 14, 2019. Recommended by Technical Editor Jamie Paik. This paper is supported by the National Research Foundation of Korea Grant funded by the Korean Government (MSIT) under Grant NRF-2016R1A5A1938472 and in part by Institute for Information and Communications and Technology (IITP) funded by Korean Government (MSIT) (No.2017-0-01778). *(Dooyoung Kim and Junghan Kwon contributed equally to this work.) (Corresponding authors: Yong-Lae Park; Sungho Jo.)*

D. Kim, S. Han, and S. Jo are with the School of Computing, Korea Advanced Institute of Science and Technology, Daejeon 305-701, South Korea (e-mail: dykim07@kaist.ac.kr; shann@kaist.ac.kr; shjo@kaist.ac.kr).

J. Kwon and Y.-L. Park are with the Department of Mechanical Engineering, Seoul National University, Seoul 08826, South Korea (e-mail: jhkwon@snu.ac.kr; ylpark@snu.ac.kr).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org provided by the authors.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMECH.2018.2874647

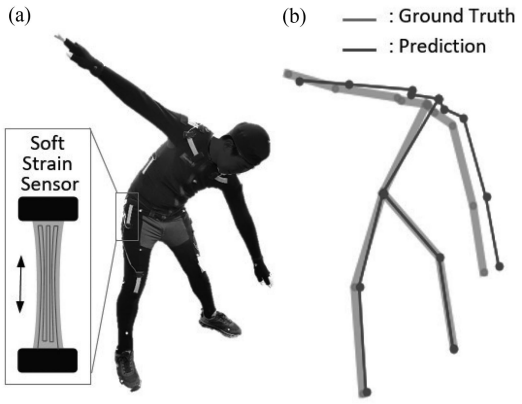


Fig. 1. (a) Prototype of the soft wearable sensing suit for full-body motion tracking. (b) Corresponding 3-D skeleton reconstructed using DFM-Net.

bration and mapping approach, three different athletic motions were tested with a human subject, and the test result showed a higher accuracy than traditional methods that were mainly based on mathematical estimation, such as LR.

The remainder of this paper is organized as follows. A brief survey of existing motion tracking and soft sensors is presented in Section II. Section III shows our soft wearable sensing suit and its signal characteristics. Section IV discusses the structure of our learning-based calibration process, followed by the experimental results that evaluates the performance of the proposed method in Section V. Finally, we conclude our research in Section VI.

II. Related Work

Body motion tracking has been a long-standing question in biomechanics and rehabilitation, and various approaches have been proposed using different types of devices, such as electrogoniometers [14], [15], camera-based optical systems [16], [17], and inertial measurement units (IMUs) [18]–[20]. In spite of their success in some level, they still have limitations. For example, electrogoniometers have difficulty in detecting multi-DOF joint motions. Although optical systems are highly useful for full-body motions with high accuracy, they are limited to only indoor use due to the multiple cameras fixed around the subject. IMU systems are able to overcome this space limitation, but they show errors with high-speed motions and position drifts for a long-term measurement. Furthermore, rigid electronics needed to be attached to different locations of the body, which may cause discomfort to the wearers.

To address the above-mentioned issues, a lower-limb soft wearable sensing suit for gait measurement has been developed using microfluidic soft strain sensors [5]. Although it showed feasibility of using soft wearable sensors, the tracking motions were limited in a sagittal plane, and the calibration was based on simple linear fitting that does not fully represent the complex body motions and the nonlinear characteristic of the soft sensors. Another approach has been made to detect 3-D motions of the ankle joint using multiple capacitive soft sensors [6]. In this case, LR was used for calibration of joint angles from multiple

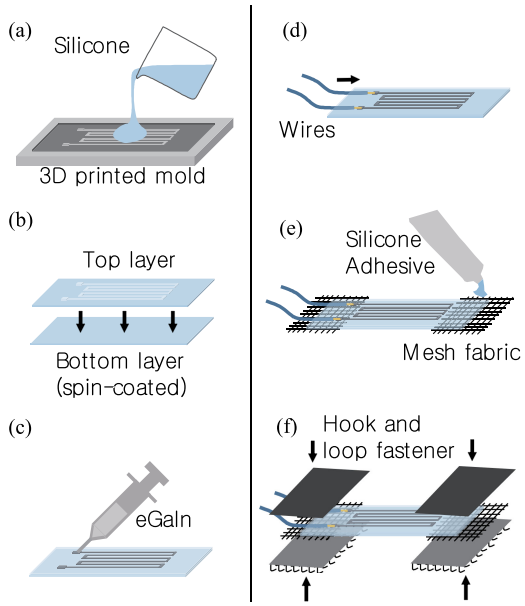


Fig. 2. Fabrication process of the soft strain sensor.

sensor signals. Although the device was easy to wear and able to detect the 3-D ankle motions, it showed installation issues on alignment, anchoring, and slippage of the soft sensors caused by deformations of the human body [10]. Soft sensors have also been used for upper body tracking [7]. Two piezoresistive soft sensors were attached to the shoulder for detecting two-DOF shoulder motions. Although it showed capability of detecting multi-DOF joint motions, the method of direct attachment to the skin may significantly reduce the practicality of the system as the number of sensors increases.

Therefore, to be practical for applications, such as rehabilitation, gaming, VR/AR, etc., the soft sensing suits need to fulfill the following conditions:

1) to cover the full range of body motion;
2) to be easy to wear and use;
3) to provide a higher accuracy than that of commercial home-entertainment products, such as Kinect [4] that has an root-mean-square error (RMSE) of 0.12 m [21].

III. Soft Wearable Sensing Suit

A. Design and Fabrication of the Soft Strain Sensor

The design of the soft sensors was based on our previous work [5], [22], and a simplified fabrication process was developed, as shown in Fig. 2.

A silicone (Ecoflex 50, Smooth-On Inc.) layer was cast using a 3-D printed mold and bonded to a spin-coated bottom silicone layer. Then, a liquid metal compound (eutectic gallium–indium; eGaIn) was injected into the microchannel, and signal wires were plugged into the microchannel directly. Next, both the ends of the sensor were reinforced with mesh fabric using silicone adhesive. Finally, hook-and-loop fasteners were directly sewn to the mesh fabric for attachment of the sensor to the full-body suit. The sewing process holds the signal wires tight and helps to prevent the signal wires from being pulled out of the sensor.

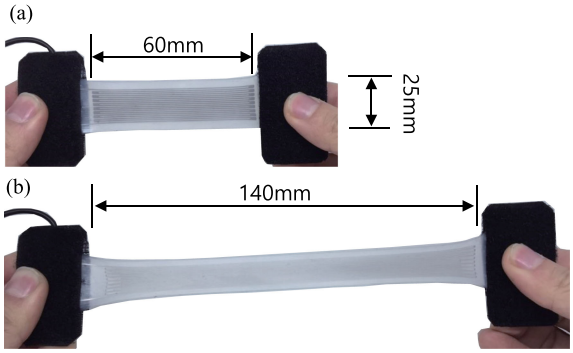


Fig. 3. Prototype of the soft strain sensor. (a) In original length. (b) In fully stretched.

TABLE I
LOCATIONS AND TARGET JOINTS OF THE ATTACHED SENSORS

Sensor ID	Location	Target joint
01, 02	Elbows	Forearms
03, 04	Top of trapezius	Upper arms and Shoulders
05, 06	Pectoralis major	Upper arms and Shoulders
07, 08	Back of deltoideus	Shoulders and spine
09, 10	Upper side of latissimus dorsi	Spine
11, 12	Lower side of latissimus dorsi	Spine
13, 14	Flanks	Spine
15, 16	Hip	Thighs
17, 18	Side of hip	Thighs
19, 20	Fore side of knee	Shins

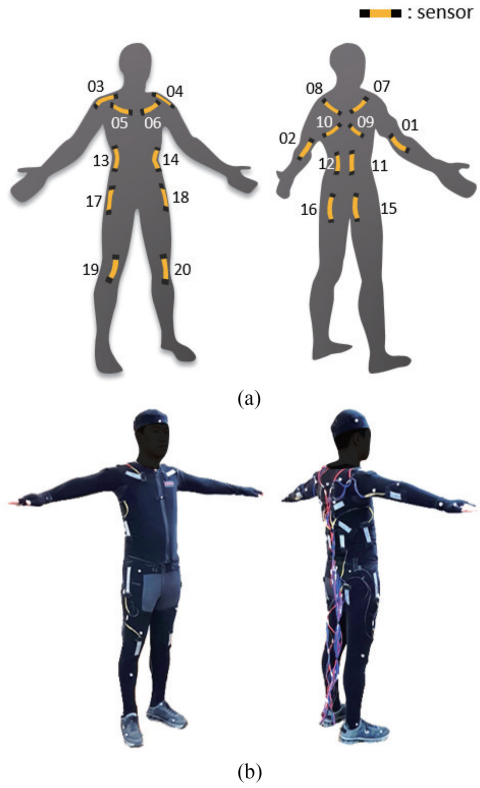


Fig. 4. (a) Locations and IDs of the sensors on the sensing suit. (b) Front and back of the prototype.

The sensor is made of only soft material that makes itself easily wearable and lightweight. It can operate up to over 130% strain due to its high stretchability (see Fig. 3). When stretched, the embedded microchannel increases electrical resistance based on the increase and the decrease of its length and the cross-section, respectively. We assume that the effect of temperature change of the human body are negligible based on the result of our previous work [23].

The resistance changes of the soft sensors are measured by a simple voltage-divider circuit [13] and a 16-bit analog-to-digital converter data acquisition module (NI USB-6259, National Instruments) [24]. The measured data is transferred to the processing computer through universal serial bus (USB) interface.

B. Sensor Placement

The location of the soft sensors were carefully selected in consideration of the position of the body joints as well as the position of the joint muscles, as shown in Table I and Fig. 4(a). A total of 20 soft strain sensors were attached on the sensing suit: 6 and 14 sensors on the lower and the upper bodies, respectively. The soft sensors were attached directly to the elbow (ID: 01, 02) and the knee (ID: 19, 20) joints that have a single DOF [5]. To measure both bending and twisting of the hip joint simultaneously, four sensors were attached to the back (ID: 15, 16) and the side (ID: 17, 18) of the hip.

In contrast to single-DOF joints, the upper body joints, especially the spine and the shoulder, make complex motions using multiple joints and muscles. Therefore, we attached multiple sensors to the upper body and predicted the motion of the spine, shoulder, and upper arms by weighted composition of the sensor outputs. Since the movements of the spine are generated by the muscles around them, we attached a total of eight sensors on these muscles. Four sensors (ID: 11, 12, 13, 14) were attached to the lower part of the upper body along the muscles at the spine, and the other four sensors (ID: 07, 08, 09, 10) were attached to the upper part of the back. In addition, the motion of the shoulder and the upper arm are more complex than the other parts due to the shoulder joints that have five DOFs: roll, pitch, yaw, and two-DOF translations in the sagittal plane. We considered the directions of the muscle fibers of the three muscles that move the shoulder: trapezius (ID: 03, 04), pectoralis major (ID: 05, 06), and back of deltoideus (ID: 07, 08).

Fig. 4(b) shows the complete prototype of the soft wearable sensing suit. As a garment base, we used a flexible spandex suit with a Velcro-friendly surface for easy attachment and detachment of the soft sensors as well as the optical tracking markers for calibration.

C. Limitations of the Sensing Suit

In order to understand the nonlinear characteristic of the soft sensor in motion, we attached a soft sensor to the knee joint [see Fig. 5(a)] and measured the sensor signals. Reference joint angle values were also collected from a motion capture system simultaneously.

Fig. 5(b) shows the measured data when the knee joint was repeatedly in flexion and extension, and (c) shows the relationship between the knee joint angles and the measured signals from the soft sensor. The result shows nonlinearity in response

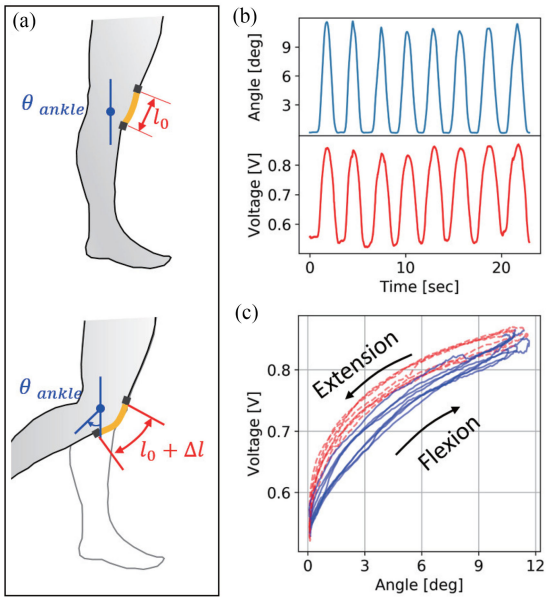


Fig. 5. Measured data from a soft sensor during eight repetitions of knee joint's flexion and extension motions. (a) Illustration of the sensor attachment on a knee joint. (b) Measured sensing data over time. (c) Relation between the knee joint angle and the measured sensor signal.

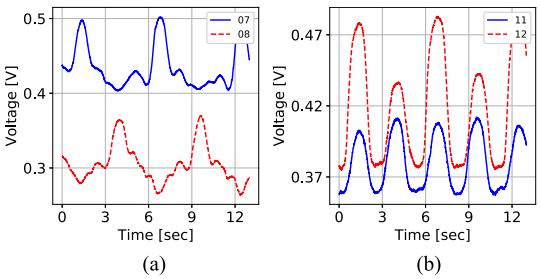


Fig. 6. Sensor outputs during a windmill motion. (a) Back of deltoideus (ID: 07, 08). (b) Lower side of latissimus dorsi (ID: 11, 12).

and hysteretic loops during flexion and extension. It was already observed that the nonlinearity was caused by the unwanted pressure when the sensor was directly placed on a bony surface [5], as well as by the hysteresis of the sensor itself [22].

Fig. 6 shows the output signals from the spine and the shoulder sensors during a windmill (WM) motion. Although the two sensors in each joint were symmetrically positioned, the signals showed difference in magnitude, noise, and pattern. This is due to the aforementioned issues of soft sensors when combined with a suit, such as alignment, anchoring, slippage of the sensors, and deformation of the human body [10], which consequently lead us to think about implementing a deep neural network to deal with the issues more effectively.

IV. Calibration Using DFM-Net

The aim of the sensor calibration is to find a calibration model F in

$$\hat{y} = F(x|\Omega) \tag{1}$$

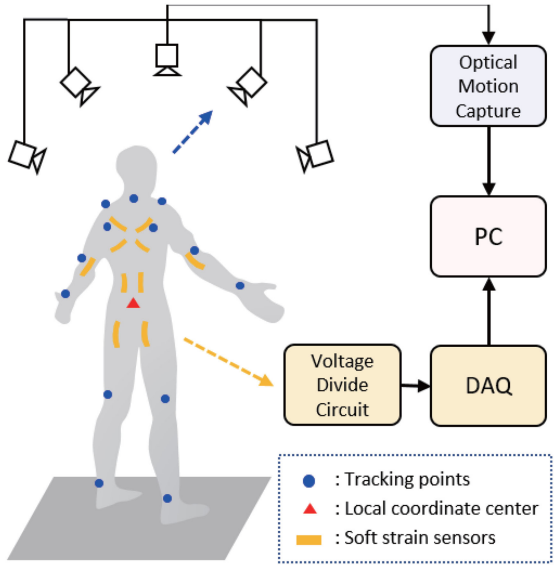


Fig. 7. Schematic representation of data acquisition setup (triangle: local coordinate center, circle: tracking points).

where Ω is the calibration parameters. The model F predicts the state $\hat{y}$ from the sensor output x. In this paper, we define a new calibration model, DFM-Net, and learn Ω through machine learning. In the training step, our model learns Ω using a training data set consisting of a pair of sensor output x and optical motion capture data y. After training, our model should be able to predict the current motion of the wearer $\hat{y}$ from x using the trained model.

A. Measurement Setup for Calibration

The environment for acquiring calibration data sets is shown in Fig. 7. To acquire training data sets, the user wore the sensing suit and stood in a 4 m $\times$ 4 m calibration space, and then made calibration motions. The resistance changes of the 20 soft sensors were then measured by the acquisition circuit and DAQ. At the same time, the reference motions were measured by an optical motion capture system (Prim13, OptiTrack), which included five optical cameras and dedicated software [16]. We selected 13 tracking points (an atlas, both scapulars, shoulders, elbows, wrists, knees, and ankles) from the reference motions to reconstruct the predicted pose skeleton [see Fig. 7(b)]. The hip center was defined as the local coordinate center of the proposed model. The data acquisition rate was set at 120 Hz for both the soft sensors and the optical motion capture system.

B. Data Preprocessing

At each time step t, data sets from the soft sensors and the motion capture system were collected as a signal vector $x^{(t)} \in \mathbb{R}^S$ and a position vector $y^{(t)} \in \mathbb{R}^{3M}$, respectively,

$$x^{(t)} = \{x_1^{(t)}, x_2^{(t)}, \ldots, x_S^{(t)}\} \tag{2}$$

$$y^{(t)} = \{y_1^{(t)}, y_2^{(t)}, \ldots, y_M^{(t)} | y_m \in \mathbb{R}^3\} \tag{3}$$

where S denotes the number of sensors, M denotes the number of tracking points, and m denotes the index of tracking points.

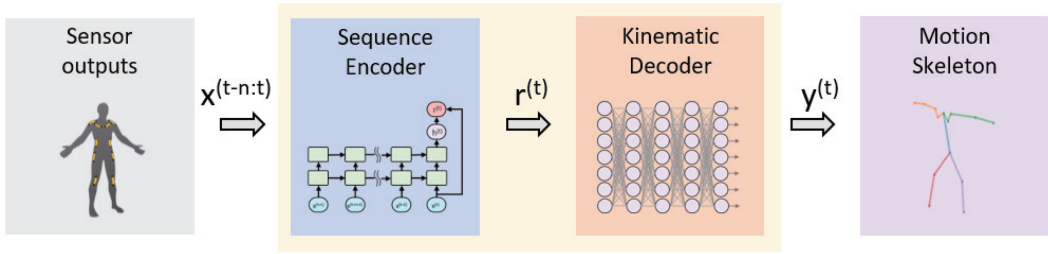


Fig. 8. Architecture of DFM-Net: $x^{(t-n:t)}$ is a sequence of the sensor output. $r^{(t)}$ is a temporal feature vector, and $\hat{y}^{(t)}$ is a predicted position of the tracking points.

In addition, a sequence of the sensor outputs $x^{(t-n:t)} \in \mathbb{R}^{n \times S}$ is used so that the sequential phenomenon of the human motion is included

$$x^{(t-n:t)} = \{x^{(t-n)}, x^{(t-n+1)}, \ldots, x^{(t)}\} \tag{4}$$

where n denotes the time window. No preprocessing procedures, such as a low-pass or band-pass filter, were used in this process.

C. DFM-Net: Deep Full-Body Motion Network

The architecture of DFM-Net for the soft wearable sensing suit is shown in Fig. 8. The model is comprised of two components: a sequence encoder network (SEN) and a kinematic decoder network (KDN). The network flow in our model is as follows. First, a feature vector $r^{(t)}$, which represents the sequential phenomenon of the sensor outputs is extracted from the input data $x^{(t-n:t)}$ using the SEN. Then, the KDN receives $r^{(t)}$ and predicts the current position of tracking points $\hat{y}^{(t)}$. In the training step, the pair of collected data sets, $x^{(t-n:t)}$ and $y^{(t)}$, are used to train the DFM-Net model. After training, the model only observes $x^{(t-n:t)}$ to predict $\hat{y}^{(t)}$. Finally, the motion skeleton is created from the predicted position at tracking points $\hat{y}^{(t)}$.

1) Sequence Encoder Network: The SEN is based on long short-term memory networks (LSTMs), which are commonly used as deep learning techniques for temporal sequence analysis [25]. The LSTMs can memorize previous inputs and use the memory to predict sequential outputs by recurrent connections in hidden units, which store temporal information. Fig. 9(a) shows an operation flow of each LSTM cell. We denote the sigmoid activation function by $\sigma(\cdot)$, the hyperbolic tangent function by $\tanh(\cdot)$, and the element-wise multiplication operation by $\odot$. At the beginning, each LSTM cell takes information as follows: the input of the LSTM cell $x^{(t)}$, the previous hidden state $h_{t-1} \in \mathbb{R}^k$, and the previous cell state $c^{(t-1)} \in \mathbb{R}^k$, where k is the dimension of the hidden and cell states. The cell first chooses what information is accepted from the previous cell state $c^{(t-1)}$ by the forget gate unit

$$f^{(t)} = \sigma\left(U_f x^{(t)} + W_f h^{(t-1)} + b_f\right) \tag{5}$$

where $U \in \mathbb{R}^{k \times k}$, $W \in \mathbb{R}^{k \times S}$, and $b \in \mathbb{R}^k$ are the input weights, recurrent weights, and biases, respectively. The next step decides what information is taken from the input vector

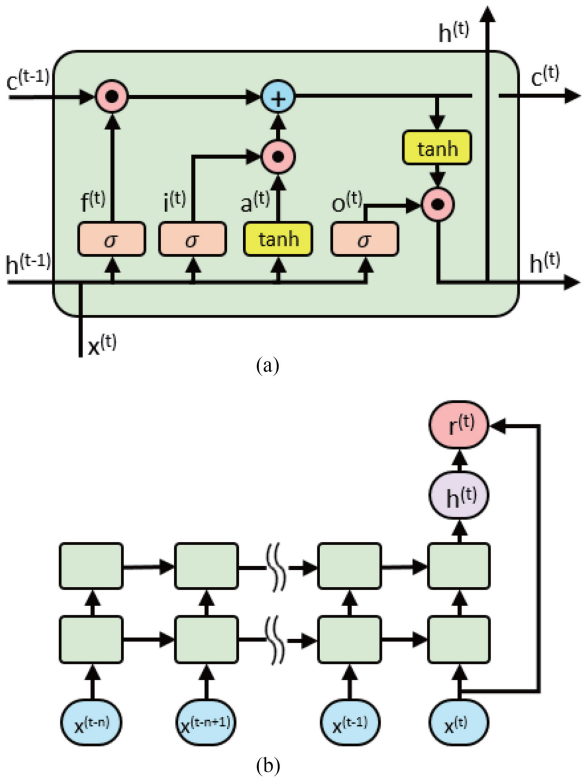


Fig. 9. (a) LSTM cell in the network model, where $\sigma(\cdot)$ is the sigmoid activation function, $\tanh(\cdot)$ is the hyperbolic tangent function, and $\odot$ is the element-wise multiplication operation. (b) Unfolded n sequential structure of the sequence encoder network.

$x^{(t)}$ by the input gate unit $i^{(t)}$ and activation unit $a^{(t)}$

$$i^{(t)} = \sigma\left(U_i x^{(t)} + W_i h^{(t-1)} + b_i\right) \tag{6}$$

$$a^{(t)} = \tanh\left(U_a x^{(t)} + W_a h^{(t-1)} + b_a\right). \tag{7}$$

Then, the cell state $c^{(t)}$ at the current time step t is derived from the above-mentioned equations as follows:

$$c^{(t)} = f^{(t)} \odot c^{(t-1)} + i^{(t)} \odot a^{(t)}. \tag{8}$$

After that, the new hidden state $h^{(t)}$ is obtained from the current cell state $c^{(t)}$ and output gate unit as follows:

$$o^{(t)} = \sigma\left(U_o x^{(t)} + W_o h^{(t-1)} + b_o\right) \tag{9}$$

$$h^{(t)} = o^{(t)} \odot \tanh(c^{(t)}). \tag{10}$$

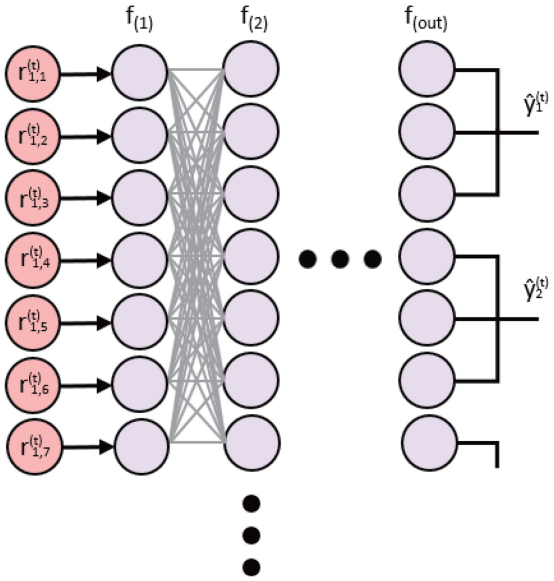


Fig. 10. Kinematic decoder network: The model predicts the location of each tracking points $\hat{y}_m^{(t)} \in \mathbb{R}^3$ from the input feature vector $r^{(t)}$.

Fig. 9(b) illustrates the unfolded n sequential structure of the SEN. In each time step t, the sequential sensor outputs $x^{(t-n:t)}$ pass through the two LSTM layers to extract the temporal information $h^{(t)}$ from $x^{(t-n:t)}$. Finally, the last sensor output $x^{(t)}$ and hidden state form the SEN $h^{(t)}$ are concatenated into a feature vector $r^{(t)} \in \mathbb{R}^{s+k}$ to represent both current and temporal features at once.

2) Kinematic Decoder Network: Now, we present a model-free approach based on a deep feedforward network to predict the position of the tracking points $y^{(t)}$ from the feature vector $r^{(t)}$ (see Fig. 10). Our network had six fully connected neural network (FCNN) layers. Each layer is modeled as in the following equation:

$$f_{(i)}(r_i; w_i, b_i) = r_i^T w_i + b_i \tag{11}$$

where r_i is an input vector of the ith layer, w_i and b_i are weights and biases, respectively. As an activation function, a rectified linear unit (ReLU) is used at each layer except the last one [26]

$$\text{ReLU}(r) = \max\{0, r\}. \tag{12}$$

From the above-mentioned two equations, we derive the ith layer output for the next layer as follows:

$$r_{i+1} = \max\{0, r_i^T w_i + b_i\}. \tag{13}$$

The output vector $\hat{y} \in \mathbb{R}^{3M}$, from the last layer $f_{(\text{out})}$, represents the 3-D coordinates of M tracking points.

3) Implementation Details: We built our model using the PyTorch deep learning framework [27]. The number of sensors S was 20, and the number of tracking points M was 13. At each time step t, we observed the past 1 s, thus the input window size n was 120 frames. The dimension of the hidden state in the SEN, k, was 128. Dropout [28] was only used in the first LSTM layer in SEN to prevent overfitting, and the dropout rate was 0.5. In KDN, the first-layer's input dimension was 148, the last-layer's output dimension was 3×13, and the others were 128 (see Table II). The Glorot initialization algorithm [29] was used to initialize the parameters in the DFM-Net model. Adam [30] was used as our optimization algorithm, and the learning rate was 0.001. As the cost function, we used a mean square error (MSE) loss defined as follows:

TABLE II
NETWORK MODELS AND PARAMETERS OF DFM-NET

Component	Layer index	Model	Parameters
SEN	1	LSTM	Hidden: 128, Dropout: 0.5
	2	LSTM	Hidden: 128
KDN	1	FCNN	In: 148, Out: 128, ReLU
	2	FCNN	In: 128, Out: 128, ReLU
	3	FCNN	In: 128, Out: 128, ReLU
	4	FCNN	In: 128, Out: 128, ReLU
	5	FCNN	In: 128, Out: 128, ReLU
	6	FCNN	In: 128, Out: 39

$$\text{MSE}(\hat{y}, y) = \frac{1}{3NM} \sum_i^N \sum_m^M \sum_d^3 (\hat{y}_{i,m,d} - y_{i,m,d})^2$$

where N is the number of samples in the test data set and d is the dimension index of the tracking points. In training, the number of epochs was 30, and the minibatch size was 500.

V. EXPERIMENTAL RESULTS

For evaluation, we captured three types of activity data sets: squat (SQ), bend and reach (BR), and WM. In each data set, the training set size was 7560 frames (63 s), and the test set size was 1560 frames (13 s). In training, the model learned a data set in which these three training sets are combined to represent various motions using one trained model.

A healthy male, 1.79 m in height, was recruited as a subject. The action space was a virtual box of 1.31 m $\times$ 1.61 m $\times$ 2.09 m (height).

We compared our method with LRs [31], which have been used for tracking 3-D ankle motions [6]. The objective of LRs is finding the coefficients $W \in \mathbb{R}^{(S\lambda+1)\times 3M}$ to minimize the residual sum of squares between the coordinates of the tracking points $Y \in \mathbb{R}^{N\times 3M}$ and the predictions XW from the sensor outputs $X \in \mathbb{R}^{N\times(S\lambda+1)}$ as follows:

$$\min_W ||XW - Y||_2^2$$

where λ is the polynomial order of the LR model. In the experimental results, we only showed a first- (LR1) and a second- (LR2) order polynomial LRs, since the calibration accuracies after third-order were lower than the second-order using the given training data sets [31]. We used the scikit-learn machine learning library to implement these comparison methods [32]. To measure the performance of the proposed method, we used a RMSE as follows:

$$\text{RMSE}(\hat{y}, y) = \sqrt{\frac{1}{3NM} \sum_i^N \sum_m^M \sum_d^3 (\hat{y}_{i,m,d} - y_{i,m,d})^2}.$$

In our experiment, our model required relatively short time to train the model and predict the position of body parts. The training process took no more than 75 s using a graphics pro-

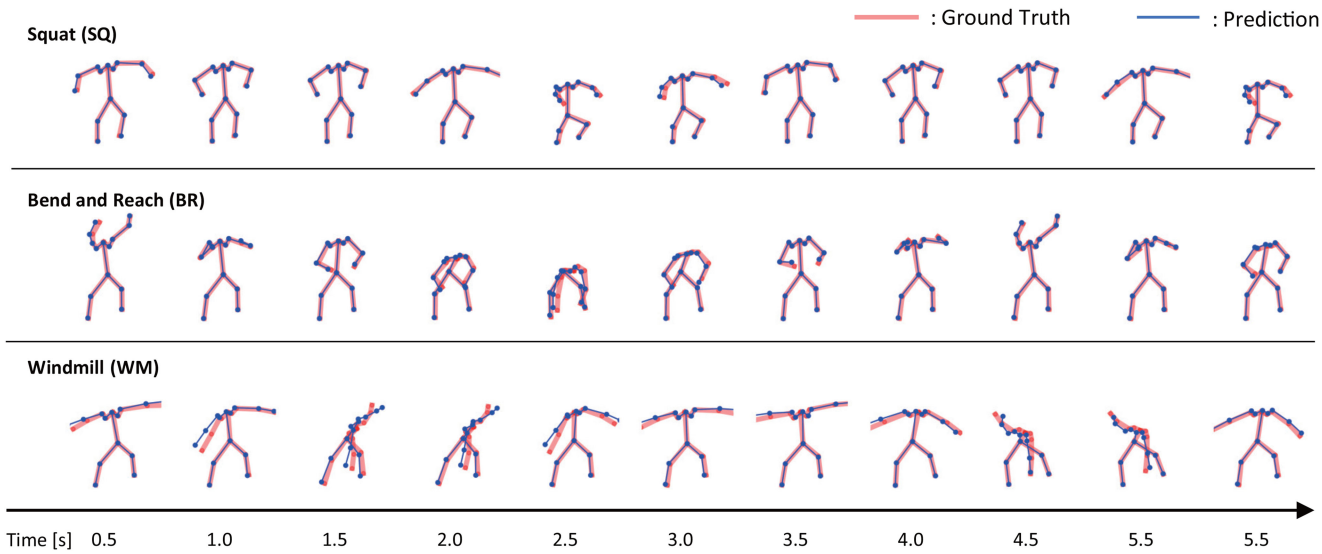


Fig. 11. Motion flow of reconstructed skeleton using the soft wearable sensing suit and the proposed calibration method, DFM-Net. The motions were captured every 60th frame (0.5 s).

TABLE III
EXPERIMENTAL RESULTS OF FULL-BODY MOTION TRACKING AND COMPARISON WITH THE OTHER CALIBRATION METHODS

RMSE (mm)	Overall	SQ	BR	WM
DFM-Net(ours)	29.5	21.9	25.2	38.7
LR1	54.6	40.9	45.0	72.4
LR2	40.0	21.4	26.8	60.3

cessing unit (GeForce GTX 1080, NVIDIA). Furthermore, the prediction took 0.21ms on an average, which is faster than the sensor data acquisition rate (120 Hz), thus enabling our method to predict the full-body motion in real time.

A. Result of Full-Body Motion Tracking

Table III demonstrates the performance of our method and the comparison with the other calibration models. The length of each training motion was 63 s, and the method was tested using four test sets: SQ, WM, BR, and a merger of all test data sets. As can be seen in the table, the overall error is only 29.5 mm, while the worst-case (WM) is just 38.7 mm. From this result, we confirm that our sensing suit and DFM-Net calibration model are accurate enough to track full-body motions, considering the action space of the human body. In comparison with the other calibration methods, DFM-Net showed the best performance in all tests except for the SQ motion. However, the difference between DFM-Net and LR2 (the second best) is negligible (0.5 mm) in the SQ. It can be inferred from this performance gap that the linear combination models have a limitation to represent the complexity of human motions.

Notably, the DFM-Net shows outstanding performance in WM (21.6 mm) compared with the other methods. The major difference between the WM and the other motions is the complexity of the spine movement. The WM has both twisting and bending of the spine, but the other motions do not have twisting. This result supports that the DFM-Net better represent the rotation of the upper body.

To evaluate the tracking quality of sequential body motions, the full-body skeletons were reconstructed from the predicted tracking points $\hat{y}^{(t)}$ every 60th frame (0.5 s). From Fig. 11, we can observe that the reconstructed skeleton follows in proximity to the ground truth from the optical motion tracking system.

In practice, the size of the calibration data set is an important factor of wearable sensors. Collecting the calibration data set is hard and time-consuming. To test how many data sets are needed to calibrate our sensing suit, we cropped each training motion data set with a certain set size and merged into one set. Fig. 12 presents the RMSE of the test result according to the size of each training motion. As seen in Fig. 12, the proposed method requires a fewer data set for the same calibration qualities compared with the other methods. Moreover, the overall RMSE of the proposed method rapidly decreased and become as small as almost 50 mm only within the first [see Fig. 12(a)]. Especially from 10 to 30 s, LR2, which showed the next best performance in overall motions, shows the worst performance due to overfitting of the noise in the model with the calibration data set [33].

The drift effect is another issue in sensor calibrations. Fig. 13 shows the changes in RMSE of the prediction results over time and their linear fitting lines. If the calibration result has a drift, the gradients of the linear fitting lines must be positive. However, they are almost zero as we can see in Fig. 13. Therefore, we can say that the drifts in the sensor signals from the suit are negligible atleast for the three test motions.

B. Detailed Observation of the Tracking Results

To present a detailed analysis of full-body motion tracking results, we categorized the tracking points into target segment sets: atlas, shoulder, elbow, wrist, knee, and ankle. Fig. 14

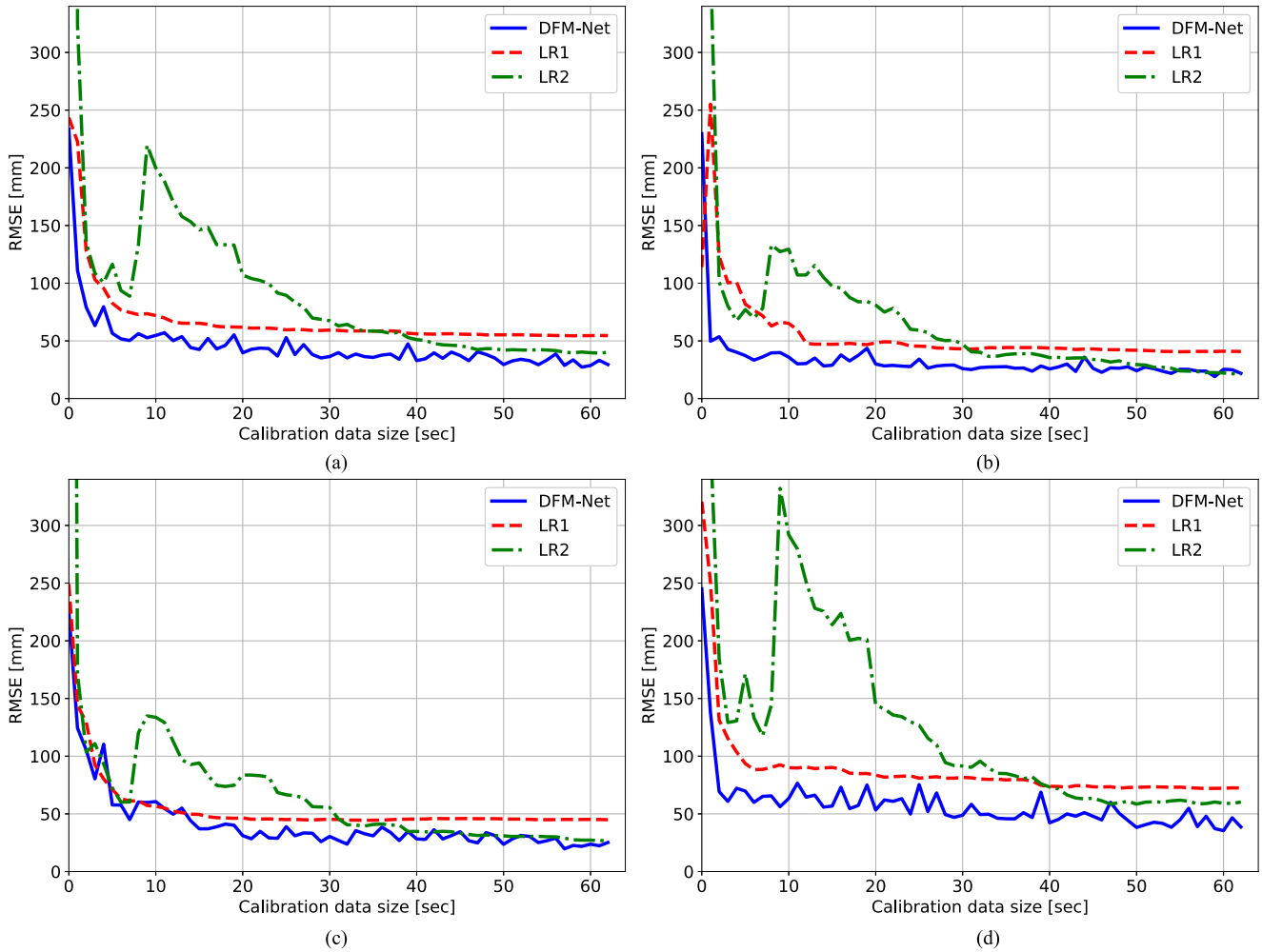


Fig. 12. Comparison of the calibration results with varied length of calibration data sets between (DFM-Net) the proposed calibration method, (LR1) first-order polynomial linear-regression, and (LR2) second-order polynomial linear-regression. (a) Overall. (b) SQ. (c) BR. (d) WM.

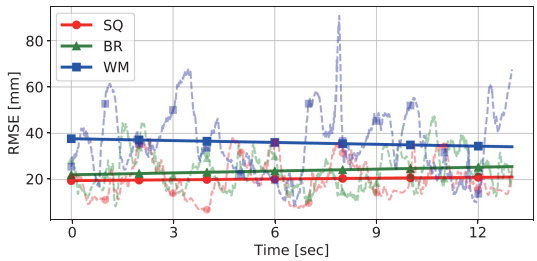


Fig. 13. The changes in RMSEs of the prediction results over the time (dashed) and their linear fitting lines (solid). SQ: $0.13\,t + 20$, BR: $0.30\,t + 22$, WM: $-0.29\,t + 38$.

illustrates the test results and Table IV summarizes the results. In Fig. 14, the cumulative distribution (y-axis) represents how many tracking points are accurately predicted within the given RMSEs (x-axis).

Fig. 14(a)–(d) illustrate the calibration quality of the complex spine motions, such as bending and twisting. The accuracy of the atlas represents the calibration quality of the spine's bending motion. In the SQ motion, the position of the atlas is just up and down. In the BR, it moves in both the upper and forward directions because the spine is bent forwards. In the WM, the side direction is added. As we can see in Fig. 14(a), only a quarter of the predictions are over the 40 mm.

The prediction quality of the twisting motions of the spine can be inferred from the RMSE of the shoulders that is shown in Fig. 14(b). The WM, other than the SQ and the BR, has the spine's twisting motion. The twisting is more complicated to measure than the bending because the sensor stretches along multiple axes. The maximum gap between the WM and the others is less than 15.2 mm, and the cumulative distribution is 0.75, as shown in Fig. 14(b). This result supports that our method can predict the spine's twisting with almost the same accuracy as that of bending.

Fig. 14(c) and (d) present the prediction result of the arm positions. In the SQ, the upper arms move from the bottom to a forward-up position. In the BR, the upper arms move up and down and the elbows are bending. In the WM, the shoulders and elbow joints are fixed. As can be seen in Fig. 14(c), which represent the prediction accuracy of the shoulder joint motion, only a quarter of the elbow predictions gave RMSEs over 40.4 mm, the worst-case motion (WM). In addition, the RMSE of the wrist is less than 29 mm except the WM motion (see Table IV).

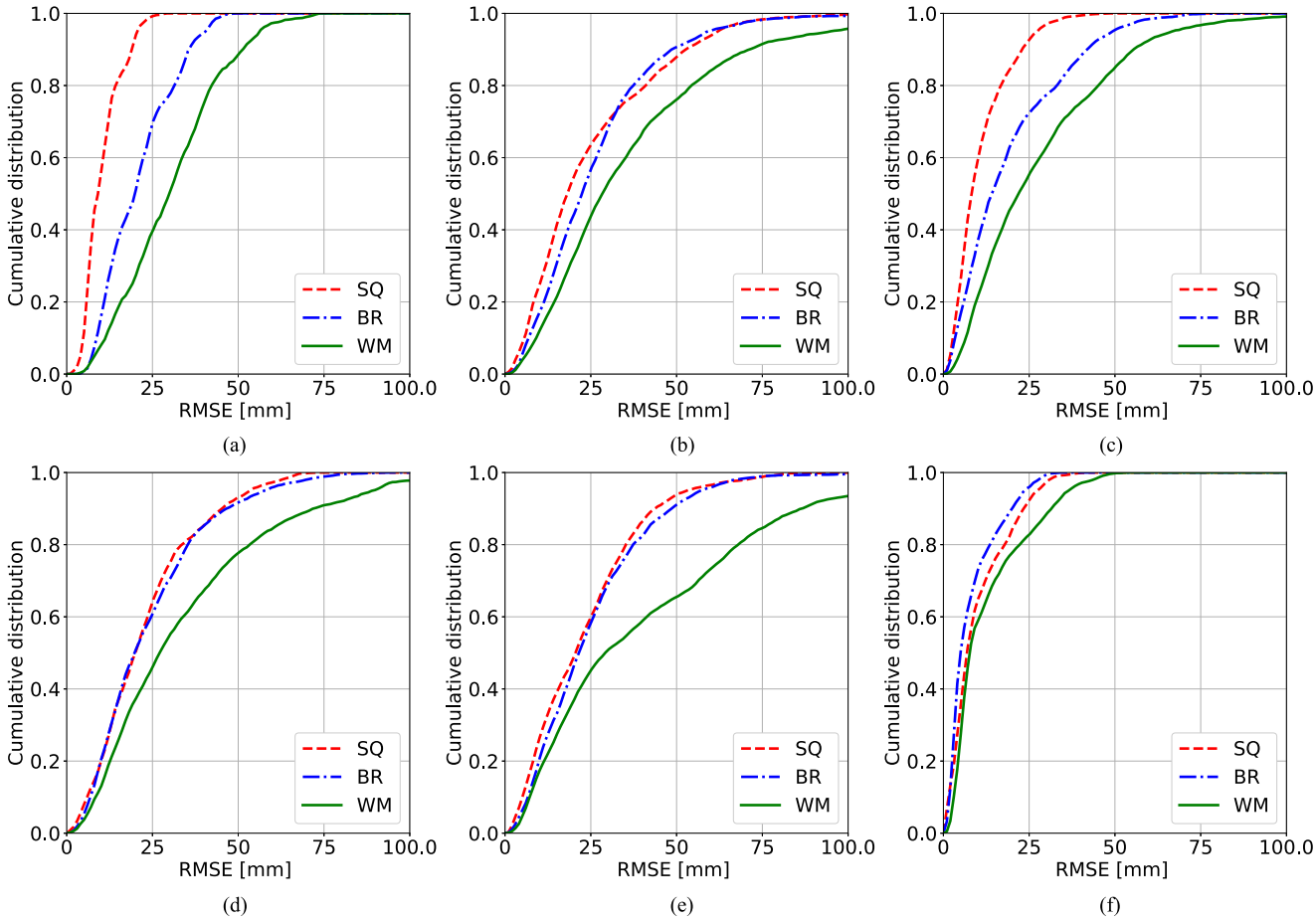


Fig. 14. Cumulative distribution of the position tracking errors. (a) Atlas. (b) Shoulder. (c) Elbow. (d) Wrist. (e) Knee. (f) Ankle.

TABLE IV
EXPERIMENTAL RESULT OF THE SELECTED SEGMENT'S POSITIONS

RMSE (mm)	Atlas	Shoulder	Elbow	Wrist	Knee	Ankle
SQ	11.6	30.9	13.5	27.3	28.2	13.1
BR	23.2	31.5	24.3	29.0	30.2	10.8
WM	33.8	47.1	34.7	43.5	53.4	17.1

Fig. 14(e) and (f) show the motions of the thighs (position of knee) and shins (position of ankle). In Fig. 14(e), the calibration quality of the twisting motion (WM) was slightly lower than the others, but only a quarter of the predictions gave RMSEs over 62.6 mm.

C. Model Analysis

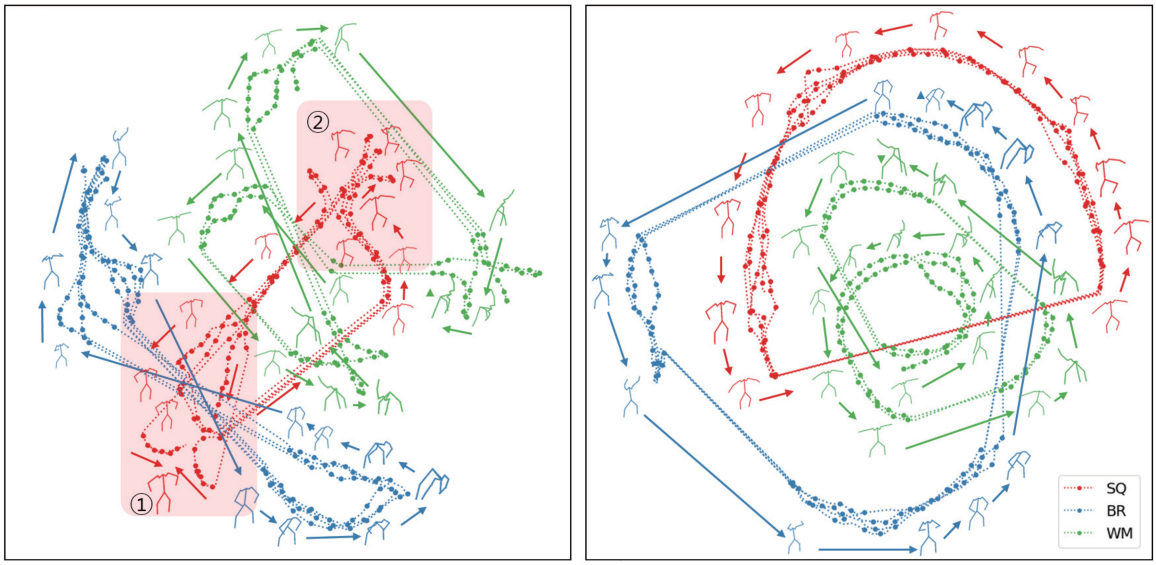
In order to understand the internal process of the DFM-Net, we compared the input feature vector x, which is the output of the suit, and the internal feature vector r, which is the output of the SEN. We projected these two feature vectors from high-dimensional space to a two-dimensional (2-D) plane by t-distributed stochastic neighbor embedding [34] to visualize and compare the characteristics of the two feature vectors. A distance between two points in the 2-D plane represents the similarity of the features at these points (less distance means high similarity).

Fig. 15 illustrates the temporal flow of the x (a), and r (b) for each motion: SQ (red), BR (blue), and WM (green). The feature flow of all motions in both x and r is in a cyclic pattern, which represents the repetition of the activity.

The feature flows in the feature vector x is highly disordered, which is likely due to the nonlinear property of the multiple soft sensors. For example, feature flows of similar SQ motions in the shaded area are highly different although they were expected to be constantly alike. On the other hand, most feature points from the motions in the shaded area are located very close to one another although they are from different sequential poses. The other two motions also show similar results.

Fig. 15(b) demonstrates that the SEN is able to extract the feature that best represents the temporal sequence and the current motion from the output signals of the soft wearable motion sensing suit. Although all three features are similar, the feature flows in feature vector r are more distinguishable than the other two. Thus, it implies that our model is more robust even with the presence of the anomalies, such as noise.

In conclusion, Fig. 15 indicates that our deep learning model successfully learned how to generalize the characteristics of the sensor data because its representational capacity was sufficient



(a) (b)

Fig. 15. Temporal flows of the features and their reference motions from the optical motion tracking system. (a) Soft sensor output features of soft wearable sensing suit x. (b) Extracted features using the SEN r from the x.

to capture the nonlinear and the hysteretic characteristics of the soft sensing suit for full-body motion tracking. Although LRs are able to calibrate the multiple sensors in our experiment, their hypothesis space were unsuitable for modeling the intractable properties due to the imperfection of its modeling capability. Thus, it was weak in the untrained anomalies, making clear the reason that our deep learning model showed better performance than the others.

VI. Conclusion

In this study, we developed a calibration method, DFM-Net, for a soft wearable sensing suit for full-body motion tracking. While previous research has mostly focused on the motion of a single joint, we measured full-body motions using multiple soft sensors. Furthermore, we defined a calibration method, DFM-Net, using deep neural networks to overcome the obstacles of the soft sensor calibration and challenges in human kinematic modeling.

To evaluate our method, three types of athletic motions were tested, and the RMSE of overall motions was 29.5 mm, while that of the worst-case motion (WM) was only 38.7 mm. The experimental results showed that the proposed method provided a higher accuracy even with a small-sized calibration data set than the other methods.

In addition, the proposed model was able to extract the features that well represent the body poses in a motion from the nonlinear raw outputs from the multiple sensors as described in the model analysis.

Our method showed the high accuracy in pose prediction, but there is a remaining challenge of a need for calibration every time when a user wears the sensing suit. One area of the future work will be development of an improved calibration model that can reuse pretrained model parameters to simplify the calibration procedure.

To the best of our knowledge, this is the first study to track the full-body motion using soft strain sensors and deep neural networks. We believe that this research will provide a new research directions in soft robotics.

Details of our implementation and source codes are available at https://github.com/KAIST-NMAIL/DFMNET.

References

[1] H. Zhou and H. Hu, "Human motion tracking for rehabilitationa survey," *Biomed. Signal Process. Control*, vol. 3, no. 1, pp. 1–18, 2008.
[2] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Image Understanding*, vol. 104, no. 2, pp. 90–126, 2006.
[3] J. C. Chan, H. Leung, J. K. Tang, and T. Komura, "A virtual reality dance training system using motion capture technology," *IEEE Trans. Learn. Technol.*, vol. 4, no. 2, pp. 187–195, Apr./Jun. 2011.
[4] Kinect for Xbox One. [Online]. Available: https://www.xbox.com/en-US/xbox-one/accessories/kinect
[5] Y. Meng *et al.*, "Wearable soft sensing suit for human gait measurement," *Int. J. Robot. Res.*, vol. 33, no. 14, pp. 1748–1764, 2014.
[6] M. Totaro *et al.*, "Soft smart garments for lower limb joint position analysis," *Sensors*, vol. 17, no. 10, 2017, Art. no. 2314.
[7] H. Lee, J. Cho, and J. Kim, "Printable skin adhesive stretch sensor for measuring multi-axis human joint angles," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2016, pp. 4975–4980.
[8] J. B. Chossat, Y. Tao, V. Duchaine, and Y.-L. Park, "Wearable soft artificial skin for hand motion detection with embedded microfluidic strain sensing," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2015, pp. 2568–2573.
[9] J. T. Muth *et al.*, "Embedded 3D printing of strain sensors within highly stretchable elastomers," *Adv. Mater.*, vol. 26, no. 36, pp. 6307–6312, 2014.
[10] C. R. Walker and I. A. Anderson, "Monitoring diver kinematics with dielectric elastomer sensors," in *Proc. Electroactive Polymer Actuators Devices, Int. Soc. Opt. Photon. 2017*, vol. 10163, 2017, p. 1016307.
[11] Y.-L. Park, D. Tepayotl-Ramirez , R. J. Wood, and C. Majidi, "Influence of cross-sectional geometry on the sensitivity and hysteresis of liquid-phase electronic pressure sensors," *Appl. Phys. Lett.*, vol. 101, no. 19, 2012, Art. no. 191904.

[12] Y.-L. Park, C. Majidi, R. Kramer, P. Bérard, and R. J. Wood, "Hyperelastic pressure sensing with a liquid-embedded elastomer," *J. Micromech. Microeng.*, vol. 20, no. 12, 2010, Art. no. 125029.
[13] S. Han, T. Kim, D. Kim, Y.-L. Park, and S. Jo, "Use of deep learning for characterization of microfluidic soft sensors," *IEEE Robot. Automat. Lett.*, vol. 3, no. 2, pp. 873–880, Apr. 2018.
[14] P. Rowe, C. Myles, S. Hillmann, and M. Hazlewood, "Validation of flexible electrogoniometry as a measure of joint kinematics," *Physiotherapy*, vol. 87, no. 9, pp. 479–488, 2001.
[15] Biometrics Ltd. [Online]. Available: http://www.biometricsltd.com
[16] Optitrack. [Online]. Available: http://optitrack.com
[17] Vicon. [Online]. Available: http://vicon.com
[18] D. T.-P. Fong and Y.-Y. Chan, "The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review," *Sensors*, vol. 10, no. 12, pp. 11 556–11 565, 2010.
[19] H. Fourati, N. Manamanni, L. Afilal, and Y. Handrich, "Complementary observer for body segments motion capturing by inertial and magnetic sensors," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 1, pp. 149–157, Feb. 2014.
[20] D. Roetenberg, H. Luinge, and P. Slycke, "Xsens MVN: Full 6DOF human motion tracking using miniature inertial sensors," Xsens Motion Technol. BV, Enschede, The Netherlands, Tech. Rep. 1, 2009.
[21] S. Choppin and J. Wheat, "The potential of the Microsoft Kinect in sports analysis and biomechanics," *Sports Technol.*, vol. 6, no. 2, pp. 78–85, 2013.
[22] Y.-L. Park, B. R. Chen, and R. J. Wood, "Design and fabrication of soft artificial skin using embedded microchannels and liquid conductors," *IEEE Sens. J.*, vol. 12, no. 8, pp. 2711–2718, Aug. 2012.
[23] D. M. Vogt, Y.-L. Park, and R. J. Wood, "Design and characterization of a soft multi-axis force sensor using embedded microfluidic channels," *IEEE Sens. J.*, vol. 13, no. 10, pp. 4056–4064, Oct. 2013.
[24] National Instruments USB-6259 DAQ. [Online]. Available: http://sine.ni.com/nips/cds/view/p/lang/en/nid/202803
[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[26] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[27] A. Paszke, S. Gross, S. Chintala, and G. Chanan, "Pytorch," 2017. [Online]. Available: https://github.com/pytorch/pytorch
[28] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[29] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[30] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Rep.*, 2015.
[31] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, vol. 821. Hoboken, NJ, USA: Wiley, 2012.
[32] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[33] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
[34] L. V. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

Dooyoung Kim received the B.S. degree in computer science from the Korea Naval Academy, Changwon, South Korea, in 2007, and the M.S. degree in computer engineering from the Seoul National University, Seoul, South Korea. Since 2015, he has been working toward the Ph.D. degree in computer science at Korea Advanced Institute of Science and Technology, Daejeon, South Korea.

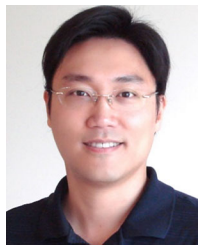
His research interests include machine learning, soft robots, and intelligent combat systems.

Junghan Kwon received the B.S. and M.S. degrees in naval architecture and ocean engineering from the Seoul National University, Seoul, South Korea, in 2008 and 2010, respectively. Since 2017, he has been working toward the Ph.D. degree in mechanical engineering at Seoul National University, Seoul, South Korea.

His research interests include soft sensors, soft actuators, soft wearable robots, and rehabilitation devices.

Seunghyun Han received the B.S. degree in computer engineering from Sogang University, Seoul, South Korea, in 2016. Since 2016, he has been working toward the M.S. degree in computer science at Korea Advanced Institute of Science and Technology, Daejeon, South Korea.

His research interests include machine learning, deep learning, and soft robots.



Yong-Lae Park (M'09) received the M.S. and Ph.D. degrees in mechanical engineering from the Stanford University, Stanford, CA, USA, in 2005 and 2010, respectively.

He is currently an Assistant Professor with the Department of Mechanical Engineering, Seoul National University (SNU), Seoul, South Korea. Prior to joining SNU, he was an Assistant Professor with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA from 2013 to 2017. His research interests include soft robots, artificial skin sensors and muscle actuators, and soft wearable robots and devices.

Sungho Jo (M'09) received the B.S. degree from the School of Mechanical and Aerospace Engineering, Seoul National University, Seoul, South Korea, in 1999, and the M.S. degree in mechanical engineering and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2001 and 2006, respectively.

From 2006 to 2007, he was a Postdoctoral Researcher with the MIT Media Lab, Cambridge, MA, USA. Since 2007, he has been with the School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, where he is currently an Associate Professor. His research interests include robotic intelligence, brain–machine interface, and wearable computing.